

Portrait: Orevaoghene Ahia

- ❖ I am a research engineer at Instadeep.
- ❖ My research is in the area of NLP, Unsupervised Learning and Transfer Learning.
- ❖ I am very passionate about work that aims to improve the performance of complex NLP architectures on low-resourced languages.

Problems with Dataset Creation for African Languages

- ❖ Most african languages are mainly spoken.
- ❖ Some language datasets are hard to annotate for instance languages with diacritics.
- ❖ Data Statements; ethical concerns.
- ❖ Different domains, sometimes training NLP models from different domains hurt performance.
- ❖ Funding is unavailable.